Augmented Reality Based on Fast Deformable 2D-3D Registration for Image Guided Surgery

Michael Scheuering[a], Christof Rezk-Salama[a],
Helmut Barfuß[b], Armin Schneider[c] and Günther Greiner[a]

[a] Computer Graphics Group, University of Erlangen-Nuremberg, 91058 Erlangen, Germany

[b] Siemens Medical Solutions, 91054 Erlangen, Germany

[c] Institute MITI, TU Munich, 81675 Munich, Germany

ABSTRACT

Augmented reality systems (ARS) allow the transparent projection of preoperative CT images onto the physicians view. A significant problem in this context is the registration between the patient and the tomographic images, especially in the case of soft tissue deformation. The basis of our ARS is a volume rendering component on standard PC platform, which allows interactive volumetric deformation as a supplement to the 3D-texture based approaches. The volume is adaptively subdivided into a hierarchy of sub-cubes, each of which is deformed linearly. In order to approximate the *Phong* illumination model, our system allows pre-calculated gradients to be deformed efficiently. The registration is realized by the introduction of a two-stage procedure. Firstly, we compute a rigid pre-registration by the use of fiducial markers in combination with an electro-magnetic navigation system. The second step accounts for the non-linear deformation. For this purpose, several views of an object are captured and compared with its corresponding synthetic renderings in an optimization method using mutual information as metric. Throughout the experiments with our approach, several tests of the rigid registration has been carried out in a real laparoscopic intervention setup as a supplement to the actual clinical routine. In order to evaluate the non-linear part of the registration, up until now several dummy objects (synthetically deformed datasets) have been successfully examined.

Keywords: Augmented reality, volume rendering, deformable model, non-linear registration, voxel similarity measure, mutual information

1. INTRODUCTION

Precise navigation within the patient's body is essential for image-guided procedures in order to assist the surgeon. One helpful possibility to ease navigation is given by augmented reality systems (ARS) that allow the overlay of navigational information like pathways or 3D CT images onto the physicians view. Especially for minimally invasive interventions those systems can be of immense importance. The bottom line of AR-systems is the registration procedure that determines the accuracy of the augmentation, adjacent to calibration of the camera and the tip transformation. In the case of rigid registration, the relation between patient body and CT mostly is calculated by the use of fiducial or anatomical landmarks.[1] Those methods suffer from the fact, that soft tissue deformation cannot be considered. In our approach, we keep track of soft tissue deformation by a 2D-3D registration of *multiple-view* video images to pre-operative 3D CT data. This is realized by matching the real video images to their synthetic views, which are generated by the use of an interactively deformable volume rendering method. Since conventional 2D-3D registration approaches mostly use surface renderings based on triangle meshes, inner volumetric structures, that directly define the shape of an object, cannot be considered

Further author information: (Send correspondence to Michael Scheuering)
E-mail: scheuering@informatik.uni-erlangen.de
Web: http://www9.informatik.uni-erlangen.de/Persons/Scheuering
Telephone: +49 9131 8529923
Address: Computer Graphics Group, Universitiy of Erlangen-Nuremberg, Am Weichselgarten 9, 91058 Erlangen, Germany

interactively. By the use of our deformable volume rendering approach, the surface to be matched to the video image is directly determined by inner structures of the CT volume with interactive frame rates, which is new in our method.

In order to give an overview of this paper, the next section presents some work, which is closely related to our approach. Section 3 introduces our volume rendering tool, that allows interactive volume deformation. Section 4 discusses the navigational components of our system including calibration procedures. In the following Section 5, we present our two tier registration algorithm. Then, Section 6 presents the results and Section 7 concludes.

2. RELATED WORK

2D-3D registration, which can principally divided into landmark-based, contour-based and intensity-based techniques, has often been used to assist physicians, when orientational aid is needed. Especially in radiological treatments like catheter navigation, 2D X-ray, portal or fluoroscopic images often have been registered to 3D CT or MR images.[2–4] Those approaches often suffer from explicit segmentation or time consuming renderings of digital reconstructed radiographs (DRR). In order to reduce the rendering time for DRRs, Goecke *et al.*[5] used accelerated direct volume rendering techniques like the shear-warp factorization.[6] Afterwards, the synthetic images from volume rendering are compared to real images by the use of intensity based metrics. A possibility to make 2D-3D registration more robust is the extension to multiple-view registration.[7,8] Here, Clarkson *et al.*[7] compare each video image with its synthetic surface renderings by evaluation of different combinations of mutual information or its derivative. Additionally, the influence of light source position on the registration result is investigated.

All those methods have in common, that they rigidly register 2D information (video, X-ray) to preoperative CT data. Soft tissue deformation is not concerned about. By the introduction of a fast deformable volume rendering method, as will be discussed in the next section, we are now capable of registering video images to pre-operative CT by the use of voxel metrics.

3. INTERACTIVE VOLUME DEFORMATION

As mentioned in the Section 1, the aim of this work is to rapidly register deformable objects to their preoperative CT volume by the use of an multiple-view approach. This requires the introduction of a new method that allows interactive deformation of volumetric datasets. Therefore, this section describes the basic concepts of our volume rendering technique that is based on general purpose graphics hardware.

In recent years efficient algorithms of volume deformation have been developed. Nevertheless, those approaches are mainly restricted to polygonal surface descriptions. Inner volumetric structures are not taken into account. To be mentioned in this context are conventional free-form modelling techniques[9] that often have been successfully applied to commercial tools. If inner structures are considered, those techniques are far away from being interactive, since they are mostly pure software solutions. Recently, general purpose hardware solutions for direct volume rendering have been published, taking into account texture mapping techniques. Especially for 3D-texture based approaches that efficiently make use of hardware assisted trilinear interpolations, recent graphics adapter releases like *NVidia GeForce3*^{TM} or *ATI Radeon*^{TM} are of immense interest, since they allow volume deformation at interactive frame rates. The approach used here is an extension to the concepts published by Rezk-Salama *et al.*[10]

Hereby, the CT volume (3D-texture space) to be deformed is divided into set of sub-cubes (*patches*), introducing a hierarchical octree structure for adaptivity (Figure 1A). The deformation is obtained by only translating the texture coordinates of each patch (Figure 1B). This leads to a flexible direct volume rendering tool, that does not require any depth sorting or slicing recalculations because of the static vertex geometry. Thus, 3D-texture space is sliced by the use of an axis-aligned method (Figure 1C), which requires additional tessellation to guarantee planar slicing polygons. Since we want to register video images to their synthetic renderings, surfaces have to be compared. In the context of volume deformation, this means that surface normals and illumination properties change as soon as the dataset is altered, according to the *Phong* illumination model. The

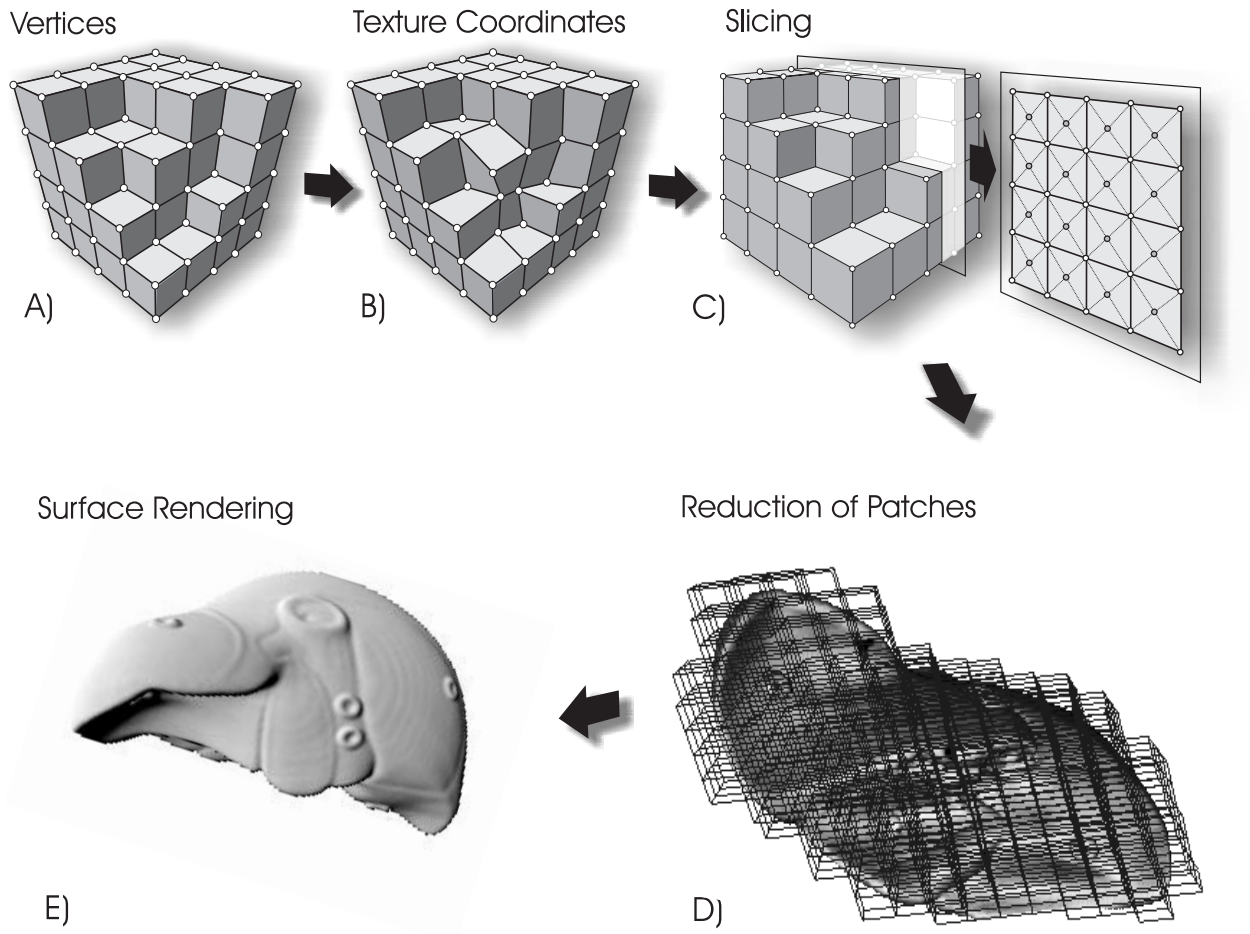

Figure 1. A) Subdivision of data: the CT volume is subdivided into a set of sub-cubes (patches) in vertex space. B) The deformation is achieved by translation of texture coordinates. C) Axis-aligned slicing of the 3D-texture space. To guarantee planar polygons in texture space, tessellation is introduced. D) Pruning the patch hierarchy: the volume rendering is accelerated by only slicing those patches that include a certain interval of grey values $\Gamma = \{i : \varrho_{min} \leq i \leq \varrho_{max}\}$. E) Result: rendering of a liver phantom.

approach described above handles this problem without any recalculations of the voxel gradient. This is done by an approximation of the trilinear mapping $\mathbf{\Phi}_t(\overrightarrow{x})$ of the elements $\overrightarrow{x}$ of a patch by an affine transformation $\mathbf{\Phi}_a(\overrightarrow{x})$[10]:

$$\Phi_t(\vec{x}) = \vec{x} + \sum_{i,j,k \in \{0,1\}} a_{ijk}(\vec{x}) \cdot \vec{t}_{ijk}, \tag{1}$$

$$\Phi_a(\vec{x}) = \mathbf{A}\vec{x} + \vec{b}. \tag{2}$$

Within that equation, a_{ijk} define the trilinear weights and $\vec{t}_{ijk}$ are the translation vectors of the patch vertices. Since for affine mappings pre-calculated gradient vectors are deformed with the transposed inverse, no recalculation is needed using Equation 3:

$$((\mathbf{A}^{-1})^T \overrightarrow{n}) \bullet \overrightarrow{l} = \overrightarrow{n} \bullet (\mathbf{A}^{-1} \overrightarrow{l}) \tag{3}$$

The variable $\overrightarrow{n}$ represents the voxel gradient and $\overrightarrow{l}$ defines the light vector. By taking into account hardware accelerated per-pixel operations for calculating the dot product in OpenGL (`GL_EXT_texture_env_dot3`), the volume can be efficiently deformed by simply converting the local light vector as in Equation 3.

Although the method described above allows fast volume deformation, a drawback is the high amount of intersection calculations per frame, since we use a decomposition into axis-aligned slices. Therefore, we introduce a pruning algorithm, that efficiently cuts those patches from the hierarchy, that do not contain any voxels within a certain user grey value interval $\Gamma = \{i : \varrho_{min} \leq i \leq \varrho_{max}\}$ (Figure 1D). Thus, an immense speedup can be achieved (Section 6). Although the pruning procedure requires a priori knowledge about the contained grey values by the user, standard *Hounsfield Scale* from CT modality can be used. The resulting rendering of a liver phantom can be seen in Figure 1E.

4. NAVIGATION AND CALIBRATION

In order to adapt the synthetic views to their related video images, one has to reproduce the pose and lens properties of the real video camera for correct video overlay in OpenGL. This requires a camera calibration procedure which determines the intrinsic and extrinsic parameters. For the intrinsic values, standard camera calibration algorithms can be used.[11] Since laparoscopic video cameras suffer from radial lens distortions, a correction algorithm is applied based on 2D-texture mapping techniques realized in OpenGL.[12] To obtain the extrinsic parameters that define the OpenGL viewing matrices for the synthetic view, an electro-magnetic tracking system (*Ascension MiniBird 800*^{TM}, 6 DOF) is used, which does not require an unfavorable intervisibility from sender to receiver like optical trackers.

For the clinical setup (minimally invasive liver interventions), we fixed the sensor ($8mm \times 8mm \times 18mm$) of the tracker on a laparoscopic camera. This setup allows a sterile plastic wrapping during intervention.[12] In order to guarantee the correct overlay, we additionally have to determine the transformation of the tip of the navigational sensor fixation to the optical center of the camera. This can be acquired by experiment or by hand-eye calibration procedures.[13]

5. REGISTRATION

In Section 3, we introduced a volume rendering module for interactive deformations of medical datasets, necessary for the generation of synthetic views. In this section, those views are now registered to their corresponding video images by the use of a two tier registration algorithm. The first step realizes a *rigid body registration* (Section 5.1), followed by a *non-linear 2D-3D registration* procedure presented in Section 5.2. In the following Section 5.3, a mathematical model that simulates soft tissue deformation to control inner volumetric structures is presented.

5.1. Rigid Registration Procedure

The basis of the two tier registration procedure is the assumption, that objects to be registered only deform non-linearly. Thus, rigid body transformations between the patient and CT can be compensated by the use of a rigid registration procedure. Therefore, we apply fiducial plastic markers, that are attached to the patients body before the CT scan. For registration purposes, we use the electro-magnetic tracking system described in Section 4 to find corresponding pairs of points within the CT and on the patient's body surface.

In order to find the rigid body transformation, we use the *Singular Value Decomposition* method (SVD)[14] as implementation. Proceeding on the assumption that objects to be registered deform non-linearly, this registration step brings us very close to the final registration result.

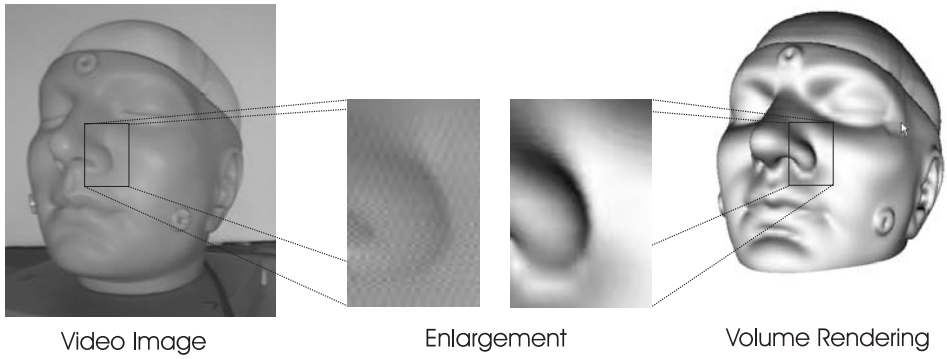

Figure 2: Similarity of grey values of an video image of plastic head phantom and its volume rendering.

5.2. Non-linear Registration Procedure

In order to realize the non-linear registration procedure, we use an intensity-based approach that directly analyzes and compares grey values of a video image with its corresponding synthetic volume rendering image. In Figure 2 a grey value video image of a plastic head phantom and its volume rendering is shown. As can been seen in the image enlargements, there exists a similarity of intensities. In our approach we use those correspondences to perform the registration. To be mentioned in this context it is of importance, that the light vectors of the real light source and of the synthetic scene (volume rendering) correspond.[7]

In recent years a lot of similarity metrics have been investigated by researchers for registration purposes. Among those metrics, *Mutual Information*[15] (MI) that statistically analyzes the information similarity, often has been successfully applied to medical image registration tasks. Given a video image V and its volume rendering R, mutual information can be calculated by the use of Equation 4:

$$MI_{rv} = \sum_{r} \sum_{v} p(r,v)(\log(\frac{p(r,v)}{p(r) \cdot p(v)})) \tag{4}$$

Thus, r and v represent the grey values of R and V, only defined in a region of overlap O. The functionals $p(r)$, $p(v)$ and $p(r,v)$ define the probability and the joint probability of r and v within the overlap O.

In order to use Equation 4 in combination with our interactive volume deformation approach (Section 3) for registration purposes, one has to optimize mutual information for a video and its rendering to reach a solution S_o:

$$S_o := argmax\{MI_{rv}\} \tag{5}$$

For implementation we use the simplex algorithm of *Powell*[14] that directly changes the coordinates of the free vertices. The degree of freedom of this optimization procedure is determined by the number of free vertices to be deformed of the volume deformation model (see Figure 1B). In Section 5.3 we introduce a higher order mathematical model that allows a *controlled* volume movement, since until now each free vertex can be changed independently from each other.

Since we want to guarantee correct registration results for more than one pair of video image and its synthetic view, we have to extend the procedure to a *multiple-view* registration method. Figure 3 shows an example of a liver phantom with different video images taken from different camera views. Clarkson *et al.*[7] presented a possible extension to multiple views for rigid registration, that combines mutual information by summation. We use this expression as optimization criterion of the non-linear deformation model (Equation 6):

$$S_m := argmax\{MI^{(i)}_{r_{(i)}v_{(i)}} + MI^{(i+1)}_{r_{(i+1)}v_{(i+1)}} + MI^{(i+2)}_{r_{(i+2)}v_{(i+2)}} + ... + MI^{(i+n=N)}_{r_{(i+n=N)}v_{(i+n=N)}}\} \tag{6}$$

Thus, Equation 6 evaluates MI for each pair $\{(V^i, R^i)|i = 1, ..., N\}$ of a video image and its synthetic view, while the combination of all pairs has to be optimized.

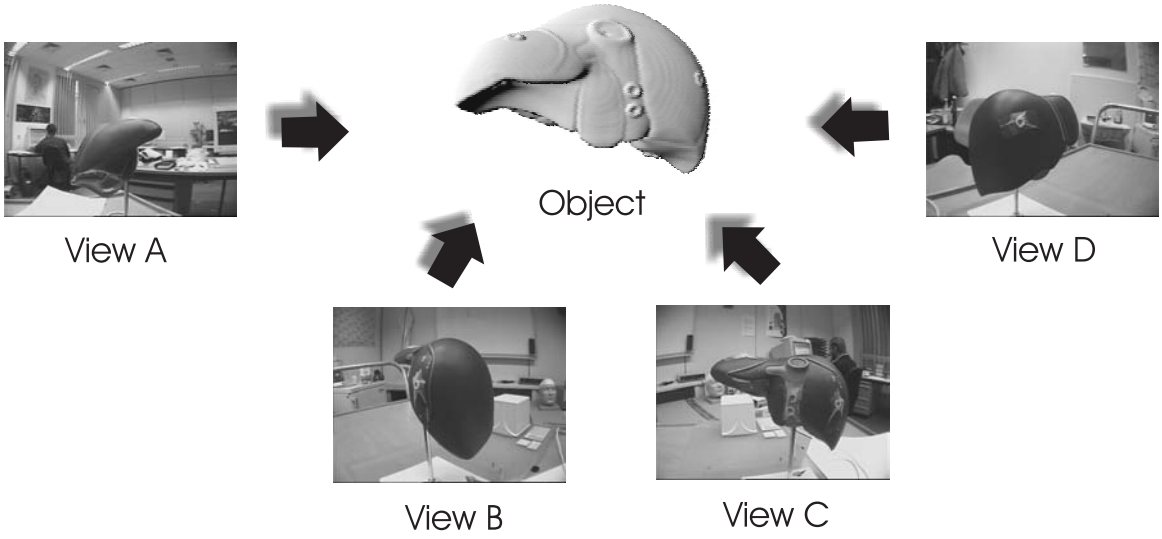

Figure 3: Multiple views of a liver phantom.

5.3. Volume control

As mentioned in the previous Section 5.2, our optimization procedure using Equation 6 allows an independent change of vertex translations to model soft tissue deformations. Although this is a very flexible possibility the degrees of freedom for optimization, determined by the free vertices defined by the user, can result in unacceptable calculation times. Additionally, independent vertex movement can hardly approximate realistic tissue deformation. In order to accelerate this process we introduce a higher order deformation model that reduces the degree of freedom and approximates soft tissue deformation. An important characteristic of the model to be chosen should be its reduced calculation time. Thus, we use a tensor product Bézier patch, because of the following reasons:

- The curve always passes through the first and the last control points, which is ideal for static vertex structure (see Figure 1A).
- The curve is always contained within the convex hull of the control points, so it never oscillates wildly away from the control points.
- The degree of freedom for optimization can drastically reduced by only deforming one or very few vertices, because the whole inner structures of the object are interpolated by the curve.

The control points of the Bézier patch are determined by the optimization procedure. The free vertices of the deformation model are then used to approximate the resulting Bézier patch.

6. RESULTS

6.1. Volume Rendering Performance

According to our interactive deformable volume rendering algorithm described in Section 3, we firstly present its performance characteristics. As previously mentioned, the approach of Rezk-Salama *et al.*[10] was modified in order to reach accelerated volume rendering times. In Figure 4 a time table is presented that shows the frames per second of a plastic head phantom ($256\times256\times32$) and a plastic liver phantom ($256\times256\times32$) of CT modality. As can be seen, an acceleration by a factor of three can be achieved (Level 3, Liver Phantom). This reduces

	Level 1 (1 Vertex)	Level 2 (27 Vertices)	Level 3 (343 Vertices)	Level 4 (3375 Vertices)	Level 5 (29791 Vertices)
Liver Phantom ($256 \times 256 \times 64$)	6.25 6.25	6.25 9.09	5.56 14.29	1.47 5.56	0.28 1.78
Head Phantom ($256 \times 256 \times 32$)	8.33 8.33	7.63 11.11	5.88 10.00	1.49 3.57	0.29 0.93

Figure 4. Rendering time in frames/s for different volume deformation levels. The numbers in brackets specify the free vertices of the volume deformation levels. The numbers to the left are the frames/s without tree pruning, those to the right with pruning technique (viewport size of 768×576).

the overall registration times immensely.* Especially for Level 5 with 29791 free vertices an improvement from 0.28 frames/s up to 1.78 frames/s was achieved.

6.2. Rigid Registration Results

In Section 5, we described the registration procedure as a two tier process. In order to clinically evaluate our rigid body registration, the algorithm was applied *intra-operatively* in *minimally invasive liver interventions* as a supplement to the clinical routine. Hereby, the liver of each patient was semi-automatically pre-segmented from the CT by a threshold based algorithm. Then, vessel or tumor structures of the liver were semi-transparently overlaid onto the laparoscopic video images by the use of volume rendering techniques in order to present navigational 3D volume information to the physician. According to the clinical setup, we used the electromagnetic tracking system described in Section 4, whereas its sensor was fixed onto the laparascope. Additionally, fiducial markers were attached to the patients body to perform the rigid registration (see Section 5.1). The clinical tests showed good orientational results for the surgeon, since additional 3D information was accurately presented intra-operatively.

6.3. Non-linear Registration Results

In order to measure the registration error for the non-linear procedure, we rigidly pre-register the dataset as presented in Section 5.1. Afterwards, the dataset is randomly deformed, serving as initial synthetic deformation of our object for non-linear registration. Since deformation is realized by vertex translation, one has to compare the distances of all free vertices of the model before and after the registration, which in general increase under deformation. Thus, our *gold standard* error ϵ_G must be zero in the optimum case. By the use of Equation 7, we are now capable of calculating registration errors by summing up all vertex distances before and after registration.

$$\epsilon = \sum_{i=0}^{N-1} (||v_b^i - v_a^i||)^2 \tag{7}$$

Hereby, v_b^i and v_a^i define the free vertices of our deformation model before and after registration, whereas N represents the number of all free vertices. Since we use Bézier curves for volume control, this measurement provides a meaningful possibility of evaluating the registration error. In Figure 5, we present two examples of this registration procedure by the use of a plastic head and a liver phantom. The column to the left shows the original video images taken from different views. The column to the middle presents the corresponding synthetic views overlaid onto video image, whereas the dataset was randomly deformed, which leads to incorrect video overlays. The last column shows registration results after the non-linear procedure.

In Figure 6 some numerical examples of the non-linear registration procedure are presented. They include the error before (ϵ_b) and after (ϵ_a) registration, the number of steps of the optimization process, the calculation

*Some volume deformation short movies can be downloaded at site:
`http://www9.informatik.uni-erlangen.de/Persons/Scheuering`.

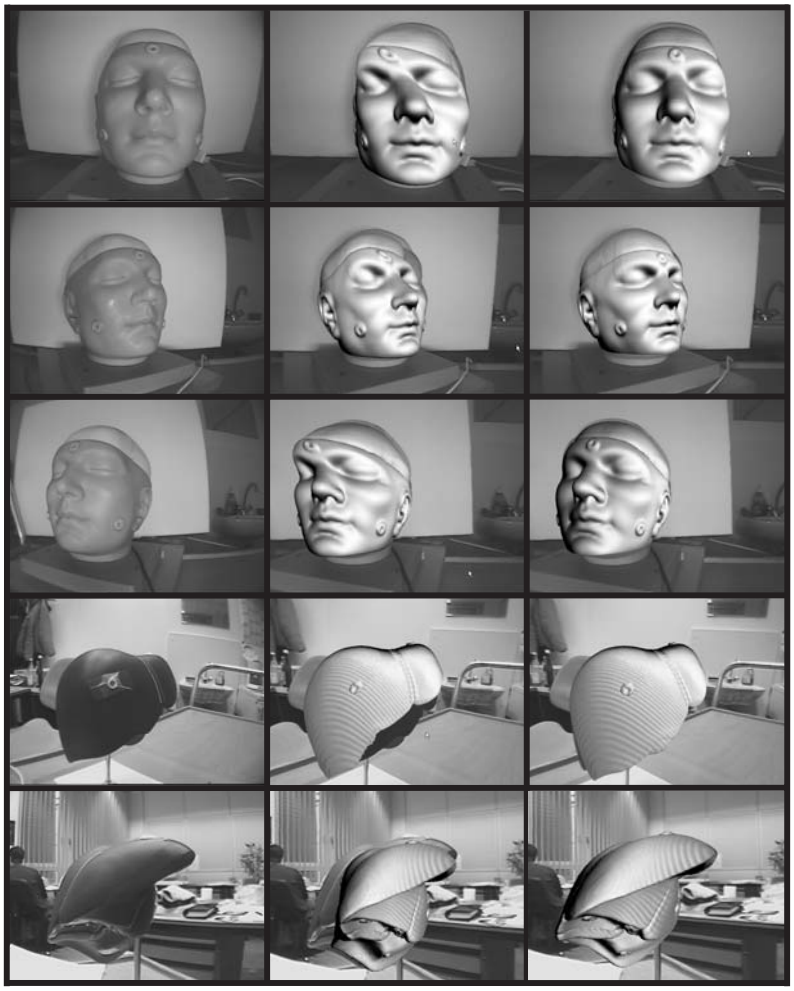

Figure 5. Multiple views of a plastic head and a liver phantom. The column to the left shows the video images taken from different views. The column to the middle shows a video overlay with synthetically deformed datasets. The right column shows the correct video overlays after registration.

time, the DOF and the number of views. One optimization step has to evaluate Equation 6 for all pairs of video image and synthetic view. As can be seen from this figure, the registration error can be drastically reduced (see Example #1) from 0.002292 up to 0.000527 within seconds with respect to the unit cube. According to the calculation times, the bulk of time is used by frame buffer copies that need approximately $301ms$ per frame. The lower images of Figure 6 show the evaluation of optimization function of Equation 6.

7. CONCLUSION AND FUTURE WORK

In this paper, we presented an two tier algorithm to perform non-linear 2D/3D registration. The algorithm is capable of registering multiple-view video images to their corresponding synthetic views, whereas the volume deforms. As presented in Section 5.1, we already evaluated the rigid image overlay in real laparoscopic interventions as a supplement to the clinical routine.

For future work, other similarity measures for robustness should be further investigated, since the registration result directly depends on the grey values of both the video image and its synthetic view. Since the rigid registration results are suited for minimally invasive procedures, the non-linear part could be successfully applied

Object	ϵ_b	ϵ_a	steps	time[s]	DOF	views
Head Phantom #1	0.002292	0.000527	55	171	3	5
Head Phantom #2	0.007934	0.005221	91	201	3	5

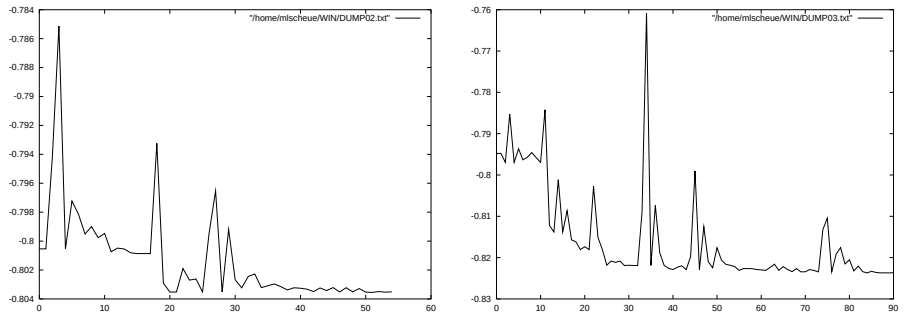

Figure 6. Result of the non-linear registration. ϵ_b and ϵ_a define the error before registration and afterwards. Additionally, optimization steps, calculation time, DOF and number of views are presented. The images below show the optimization functional (Equation 6) of Head Phantom #1 and #2. The datasets have the size presented in Fig. 4.

for open liver interventions, because multiple video images from different views can be easily taken. Summing up, our approach to augment the reality for image guided procedures provides a powerful and flexible possibility to present 3D information in order to assist the surgeon intra-operatively.

ACKNOWLEDGMENTS

I like to thank Siemens Medical Solutions Forchheim/Bavaria for the CT scans and G. Soza for fruitful scientific discussions.

REFERENCES

1. K. Masamune, Y. Masutani, S. Nakajima, I. Sakuma, T. Dohi, H. Iseki, and K. Takakura, "Three-Dimensional Slice Image Overlay System with Accurate Depth Perception for Surgery," in *Proc. of MICCAI 2000*, pp. 395–402, 2000.
2. R. Bansal, L. H. Staib, A. Rangarajan, J. Knisely, R. Nath, and J. S. Duncan, "Entropy-Based, Multiple-Portal-to-3DCT Registration for Prostate Radiotherapy Using Iteratively Estimated Segmenation," in *Proc. of MICCAI 1999*, pp. 567–578, 1999.
3. Y. Kita, D. L. Wilson, and A. Noble, "Real-time Registration of 3D Cerebral Vessels to X-ray Angiograms," in *Proc. of MICCAI 1999*, pp. 1125–1133, 1999.
4. R. Grzeszczuk, S. Chin, R. Fahrig, H. Abassi, H. Holz, D. Kim, J. Adler, and R. Shahidi, "A Fluoroscopic X-Ray Registration Process for Three-Dimensional Surgical Navigation," in *Proc. of MICCAI 2000*, pp. 551–556, 2000.
5. R. Goecke, J. Weese, and H. Schumann, "Fast Volume Rendering Methods for Voxel-based 2D/3 Registration - A Comparative Study," *Proceedings of International Workshop on Biomedical Image Registration '99* (89–102), 1999.
6. P. Lacroute and M. Levoy, "Fast Volume Rendering Using a Shear-Warp Factorization of the Viewing Transformation," *Comp. Graphics* **28**(4), 1994.
7. M. J. Clarkson, D. Rueckert, D. L. Hill, and D. J. Hawkes, "A multiple 2D video - 3D medical image registration algorithm," in *Proc. of SPIE Medical Imaging 2000*, 2000.
8. W. M. W. I. M. E. Leventon and W. E. L. Grimson, "Multiple View 2D-3D Mutual Information Registration," in *Image Understanding Workshop*, 1997.
9. D. Bechmann, "Space Deformation Models Survey," in *Computers & Graphics*, pp. 571–586, 1994.

10. C. Rezk-Salama, M. Scheuering, G. Soza, and G. Greiner, "Fast Volumetric Deformation on General Purpose Hardware," in *Proc. of SIGGRAPH/Eurographics Workshop on Graphics Hardware*, 2001.
11. R. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3-D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," in *Radiometry – (Physics-Based Vision)*, L. Wolff, S. Shafer, and G. Healey, eds., Jones and Bartlett, 1992.
12. M. Scheuering, C. Rezk-Salama, H. Barfuss, K. Barth, A. Schneider, G. Greiner, G. Wessels, and H. Feussner, "Intra-operative Augmented Reality With Magnetic Navigation And Multi-texture Based Volume Rendering For Minimally Invasive Surgery," in *Rechner- und Sensorgestuetzte Chirurgie*, pp. 83–91, 2001.
13. R. Horaud and F. Dornaika, "Hand-Eye Calibration," *IJRR* **14**, pp. 195–210, June 1995.
14. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge (UK) and New York, 1st ed., 1986.
15. P. Viola and W. M. W. III, "Alignment by maximization of mutual information," *International Journal of Computer Vision* **24**(2), pp. 137–154, 1997.